Georgia Tech

DYNAMIC DIFFERENTIAL LOCATION PRIVACY WITH PERSONALIZED ERROR BOUNDS

LEI YU, LING LIU AND CALTON PU

COLLEGE OF COMPUTING

GEORGIA INSTITUTE OF TECHNOLOGY

CREATING THE NEXT®

Location based services and Privacy issues

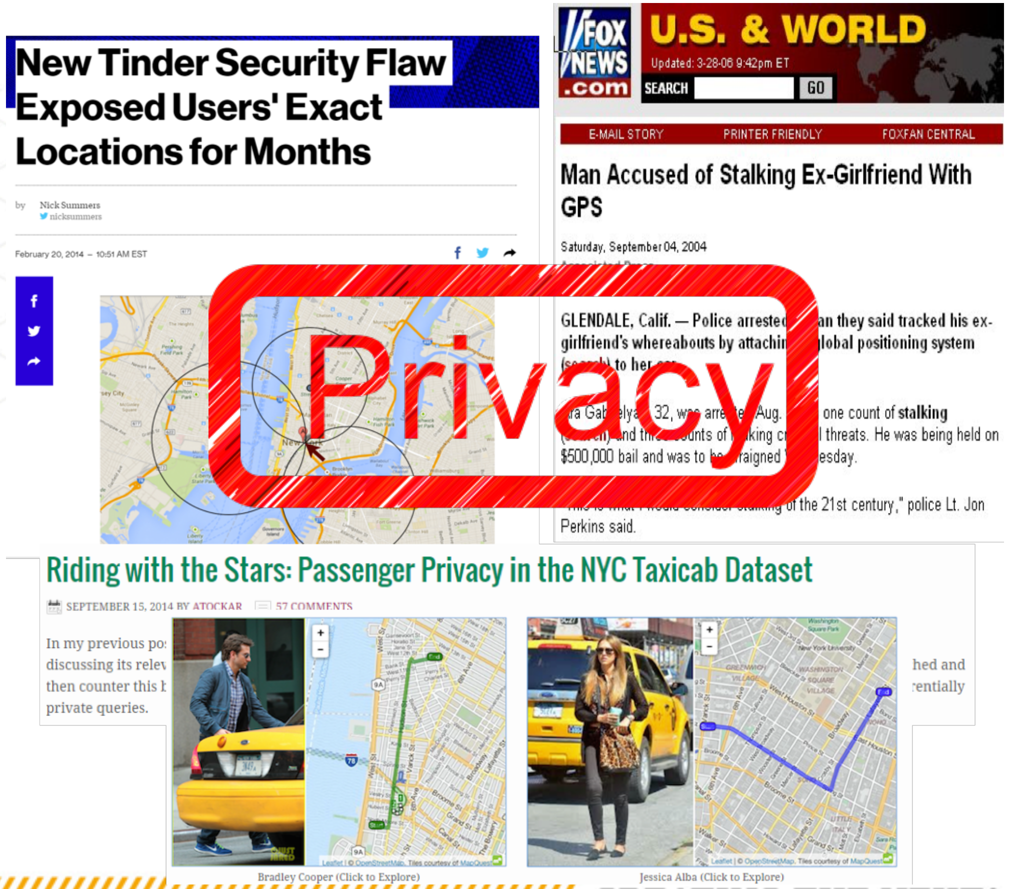
Marketing

Social Networking

Location based services

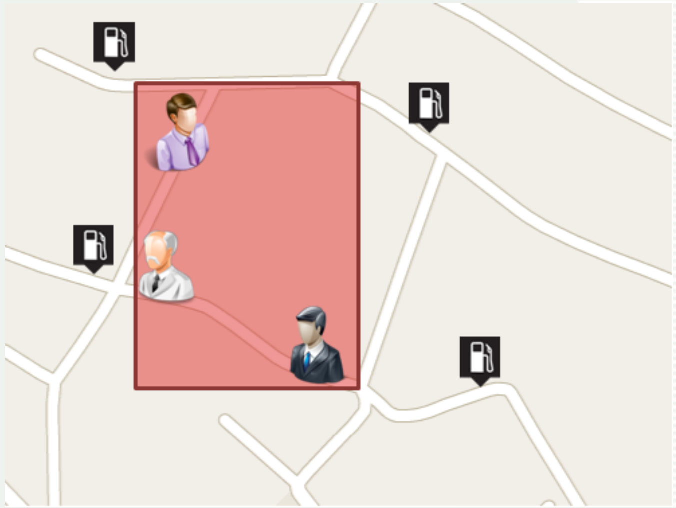
Gaming

Sports, Navigation, Health, Media, etc….



Location Privacy Protection

Anonymization

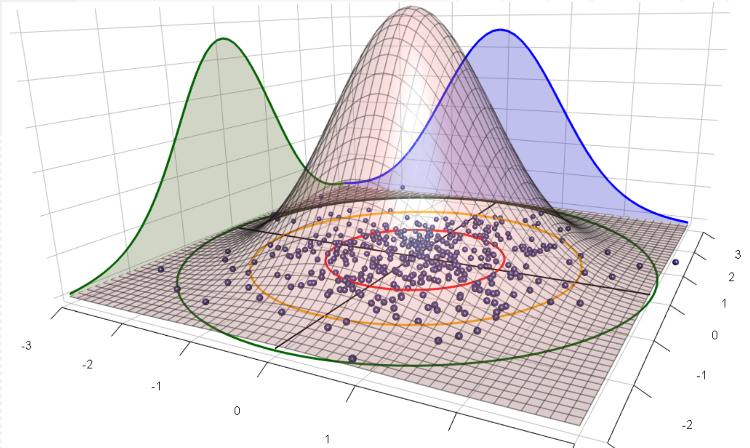


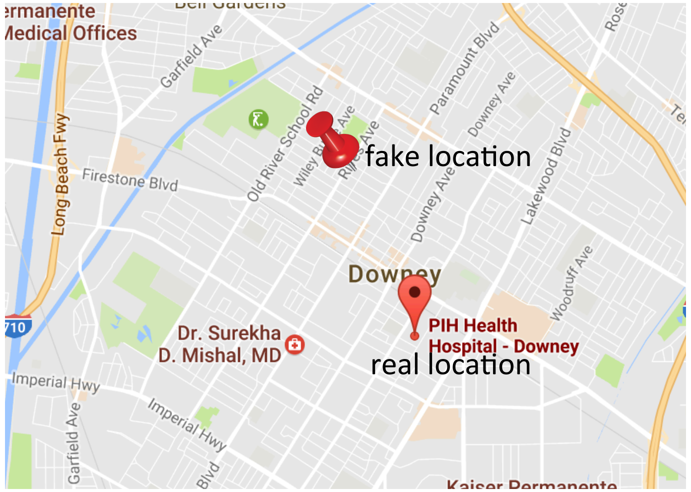
- *K*-anonymity
- trusted third-party anonymization server

Location Privacy Protection

Location Obfuscation

- Use a fake location instead of the true location
- User-centric
- Client-side





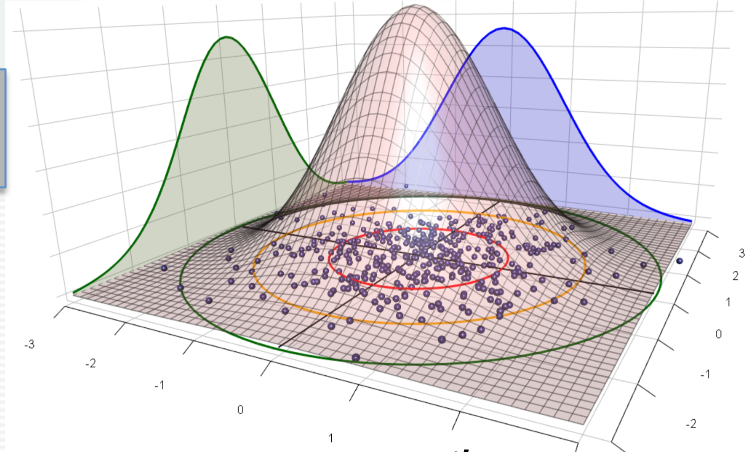
$$p(x' \mid x) = \Pr(pseudo\_location = x' \mid actual\_location = x)$$

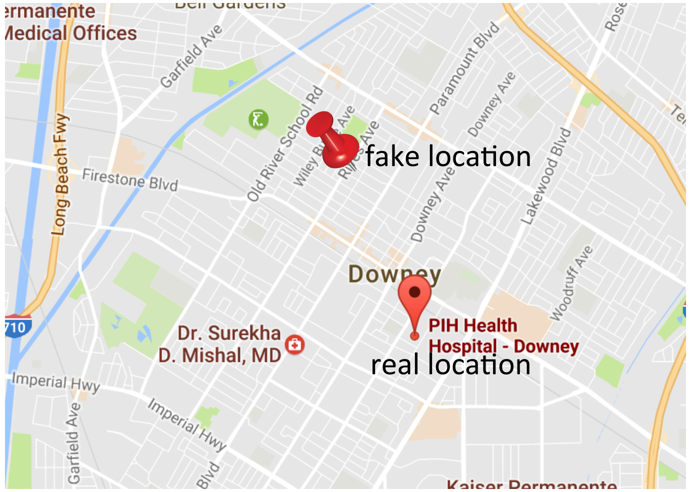
Location Obfuscation

Privacy Notion

Randomized mechanism

Utility





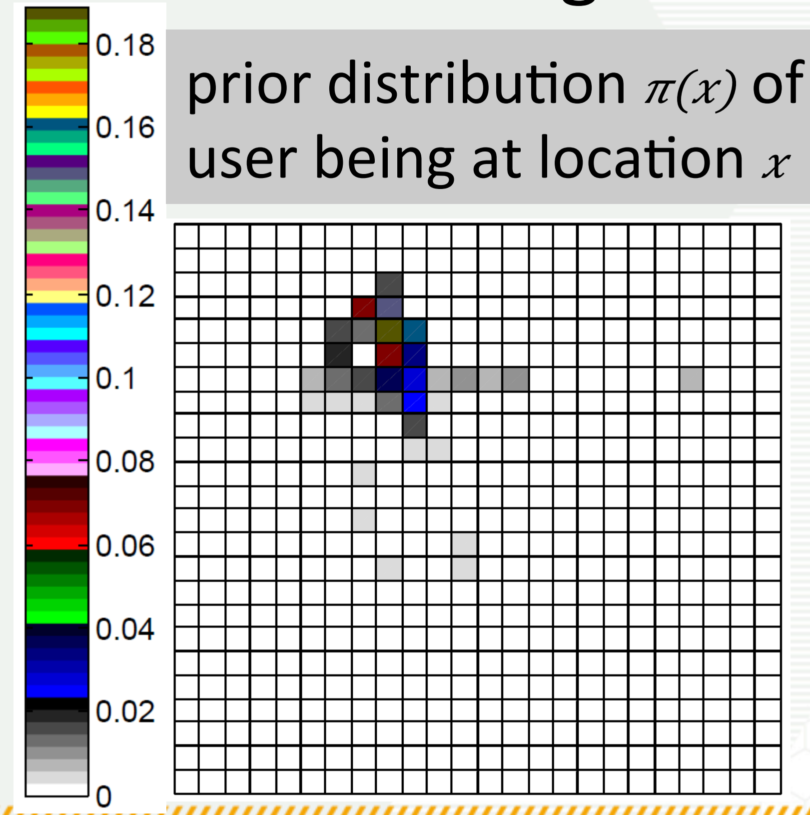
$$p(x' \mid x) = \Pr(pseudo\_location = x' \mid actual\_location = x)$$

Existing Techniques

- Privacy Notions:
  - **Expected inference error**
  - **Geo-indistinguishability**

Expected inference error

- The expected distance between the user's real location and the location guessed by the adversary.



prior distribution $\pi(x)$ of the user being at location x

Given observation x', the probability of actual location being x

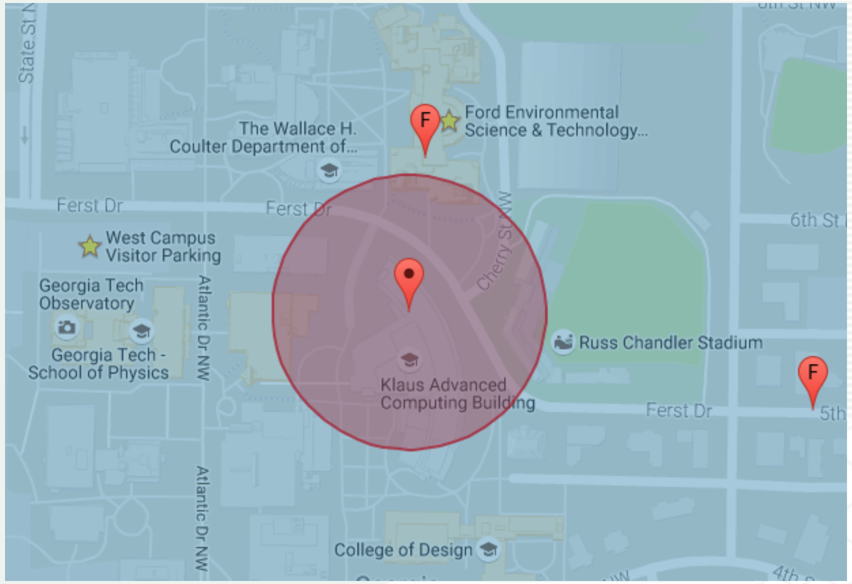
$$\Pr(x|x') = \pi(x)f(x'|x)/\sum_{x\in\chi}\pi(x)f(x'|x)$$

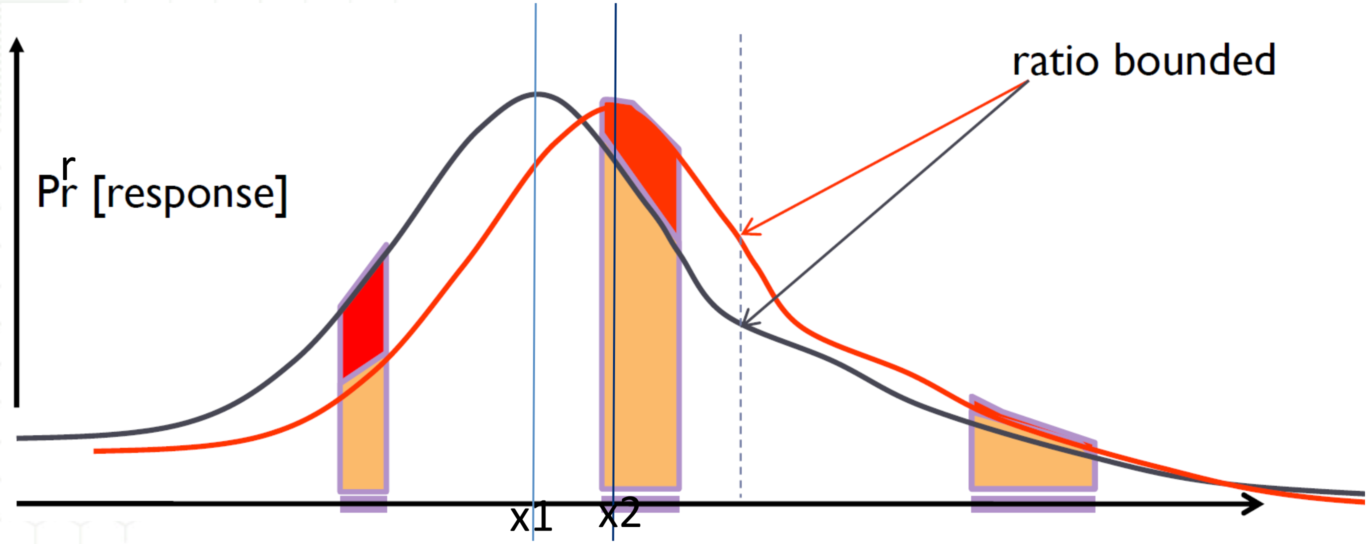
Geo-indistinguishability

Georgia Tech

For any two points *x*, *y* in the protection circular area of radius *r* centered at the actual location, by $\epsilon_g = \epsilon/2r$

$$f(x' \mid x)/f(x' \mid y) \leq e^{\epsilon}$$





Existing Techniques

Georgia Tech

- Privacy Notions:

Expected inference error	Geo-indistinguishability
Bayesian inference	differential privacy
Rely on a specific prior distribution of user's real location	only depends on the mechanism and does not depend on any prior
Not robust against any other prior distribution	Adding noise regardless of any prior can be inefficient and insufficient for privacy protection

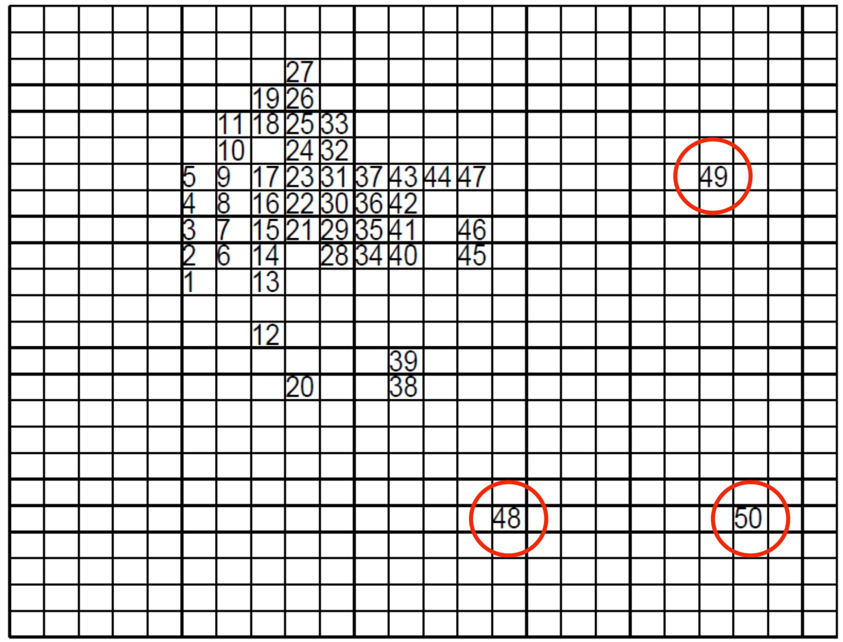
Our work

- Limitation of Geo-indistinguishability
- Two-phase location obfuscation framework
  - Adaptive noise level for different locations with guaranteeing a minimum level of inference error
  - Customizability
    - Instantly specify his privacy preference for his current location
    - Existing mechanisms are computed statically once for all, and cannot efficiently support customizability

Experimental Illustration

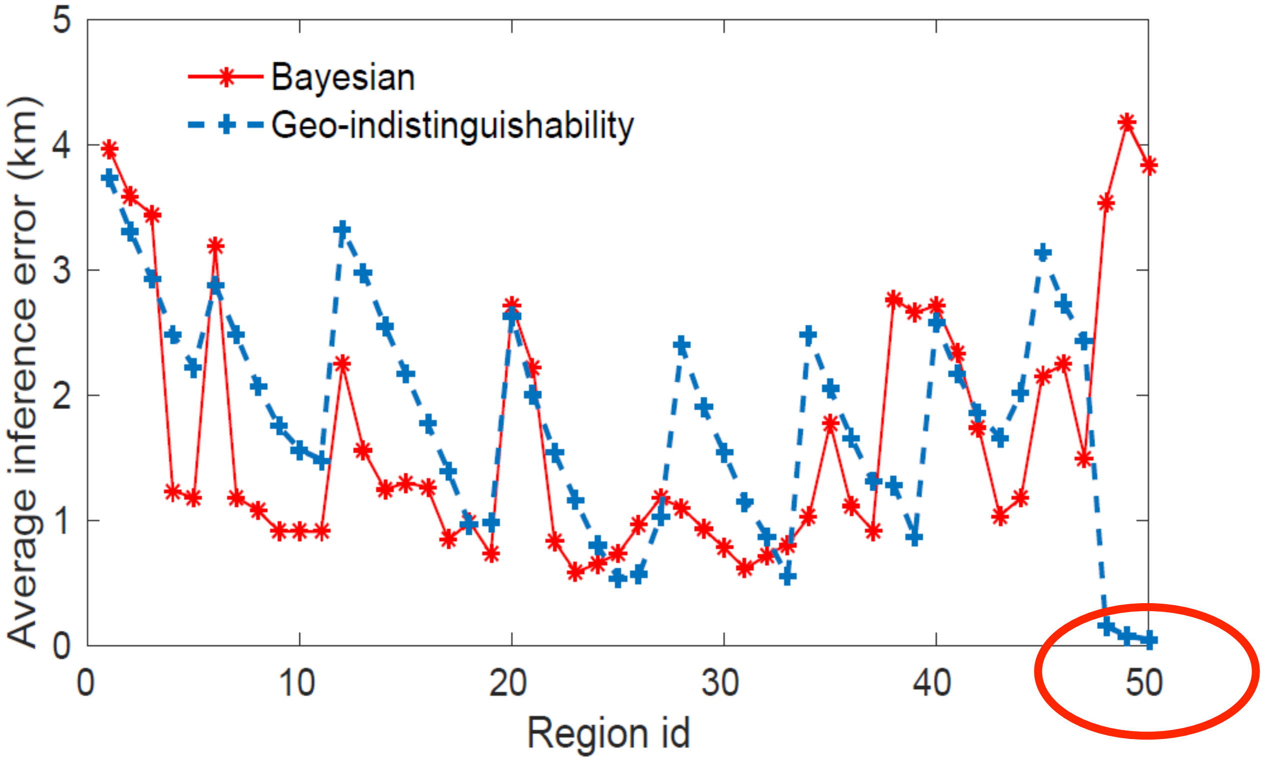
- Existing mechanisms
  - Optimal Bayesian mechanism [R. Shokri et al., 2012]
  - Optimal geo-indistinguishable mechanism [N. E. Bordenabe et al., 2014]

Experimental Illustration



50 regions with prior probability >0

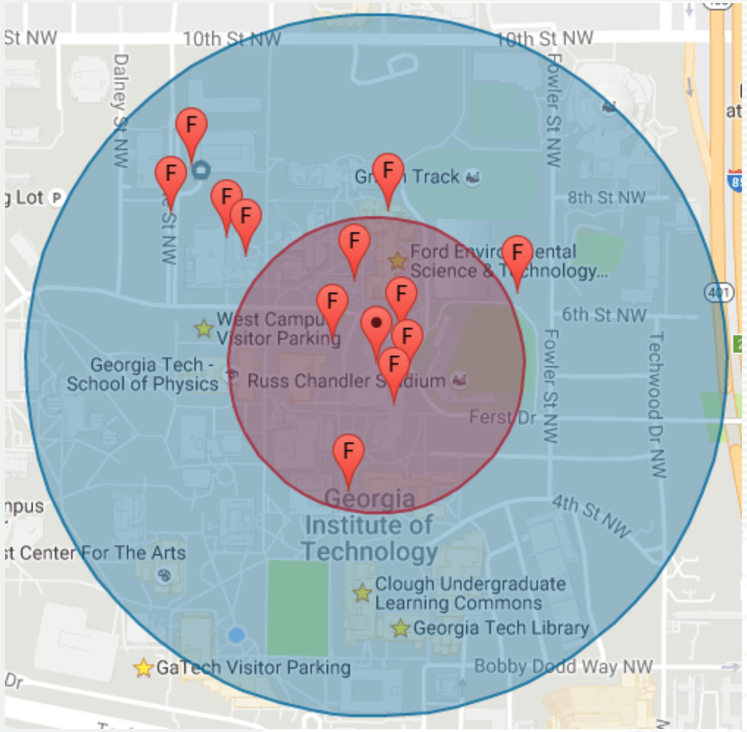
Dataset: GeoLife GPS Trajectories dataset
Formatted as in [N. E. Bordenabe et al., 2014]

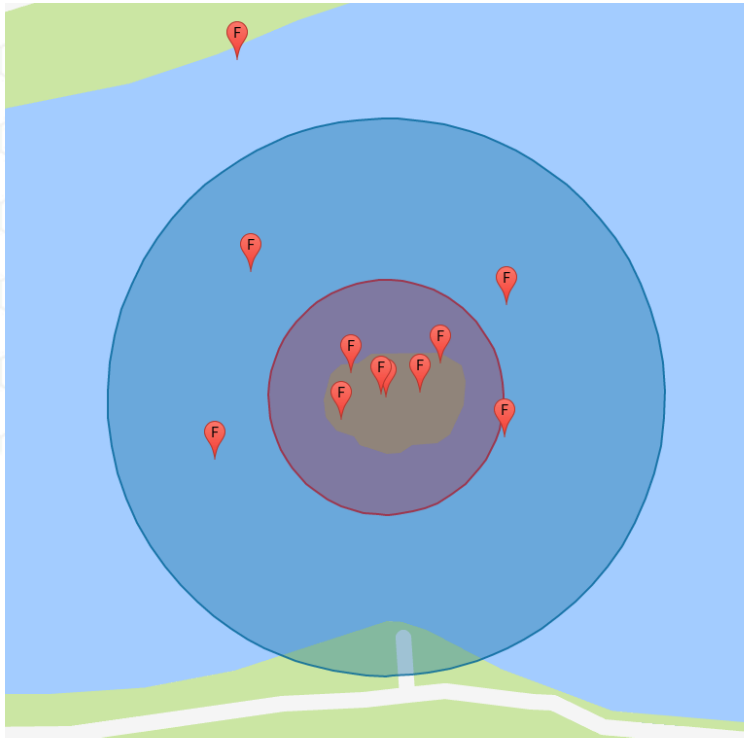


Two mechanisms that achieve the same location privacy in terms of overall expected inference error weighted by prior probability

Experimental Illustration

- Geo-indistinguishability



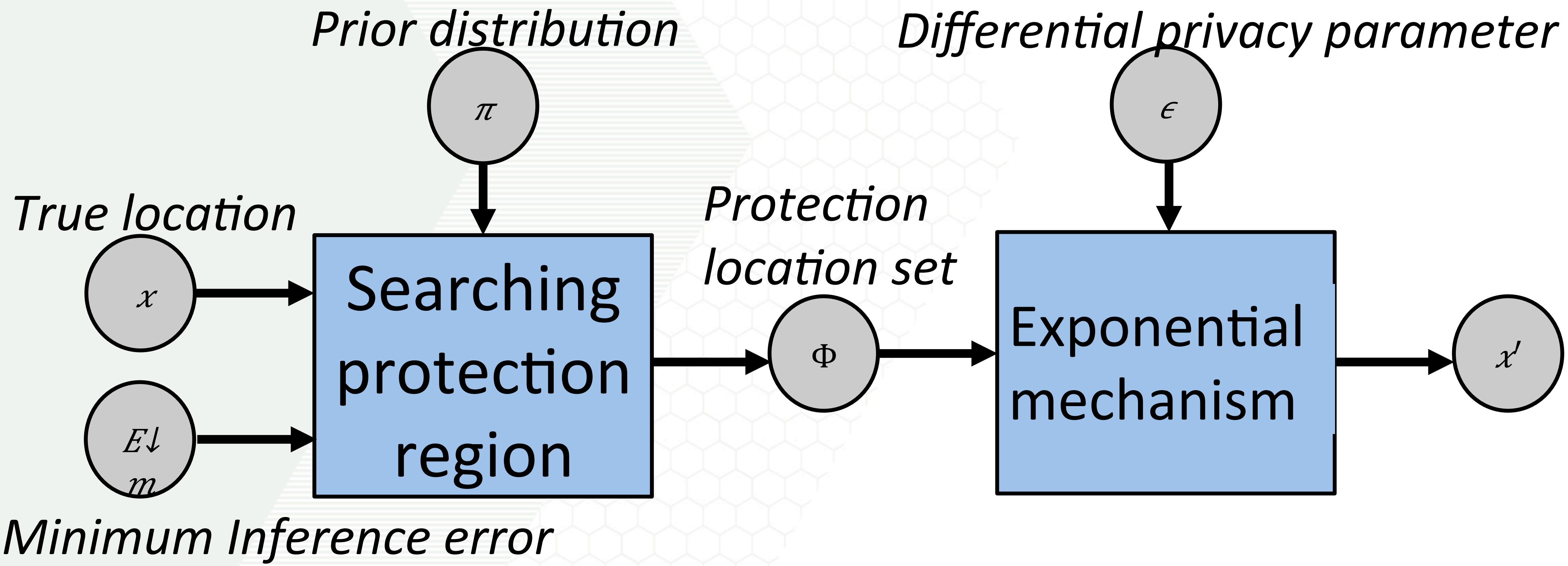


Planar Laplacian Mechanism, Pr(pseudo-location in blue circle) $\geq 95\%$

Not Adaptable: Uniform noise level either insufficient location protection at some skewed locations in terms of prior information or excessive noise for protection at other locations

Two-phase framework

- Combine expected inference error and Geo-indistinguishability



Relationship between two privacy notions

- Geo-indistinguishability
  - Any two locations x, y in the protection region Φ, $f(x' \mid x)/f(x' \mid y) \le e^{\epsilon}$

- Lower bound of conditional expected inference error

$$\min_{\hat{x}} \sum_{x \in \chi} \Pr(x \mid x')\, d(\hat{x}, x) \ge e^{-\epsilon} \min_{\hat{x}} \sum_{x \in \Phi} \frac{\pi(x)}{\sum_{y \in \Phi} \pi(y)}\, d(\hat{x}, x)$$

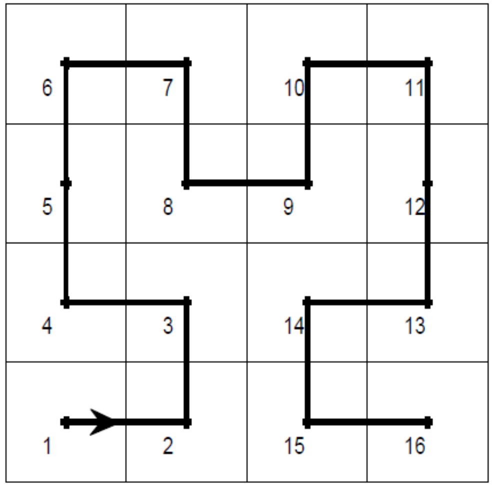
Protection Location Set

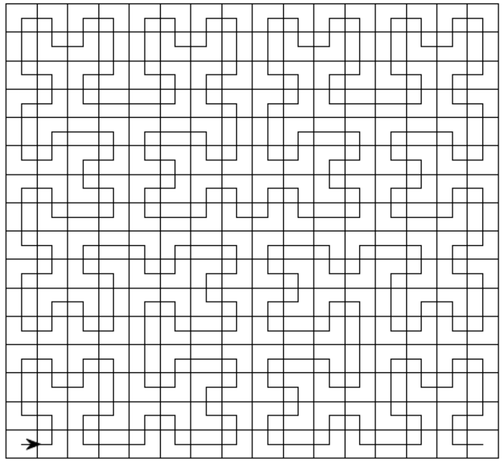
- *Theorem*: *For a location obfuscation mechanism that achieves ϵ-differential privacy on protection location set Φ, if E(Φ) $\geq e^{\epsilon} E_{m}$, the optimal inference attack using any observed pseudo-location x', the expected inference error $\geq E_{m}$.*

$$E(\Phi) = \min_{\hat{x}} \sum_{x \in \Phi} \pi(x) / \sum_{y \in \Phi} \pi(y) \; d(x, \hat{x})$$

Phase I: Search Protection Region

- $E(\Phi) \geq e^{\epsilon} E_m$
- Hilbert-curve based searching
  - Larger diameter of protection location set indicates higher noise level
  - Improvement with multiple rotated Hilbert curves





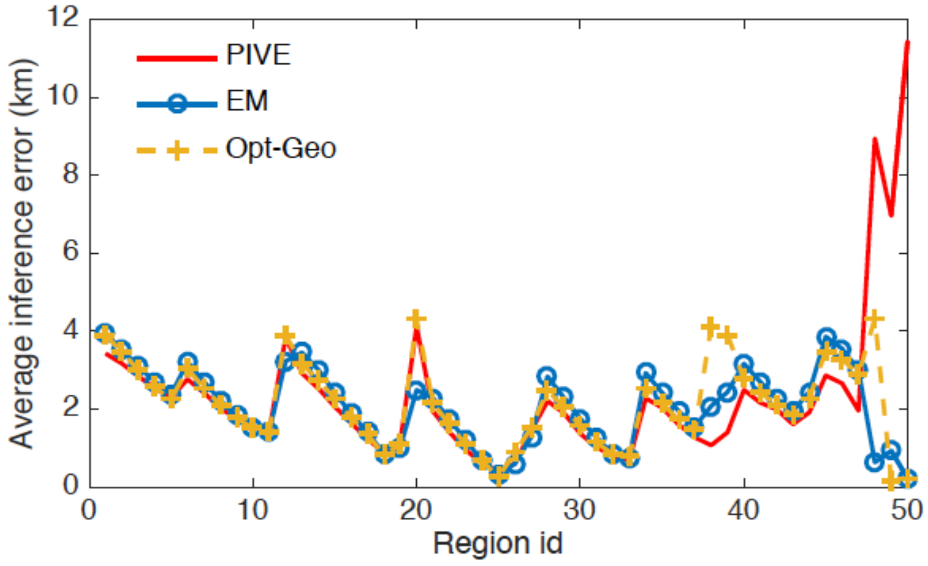
Phase II: Exponential mechanism

- Given the user's location x and location protection set Φ, the exponential mechanism selects and output a pseudo-location x' with probability proportional to $\exp(-\epsilon d(x,x')/2D)$, where D is the diameter of Φ.

Evaluation

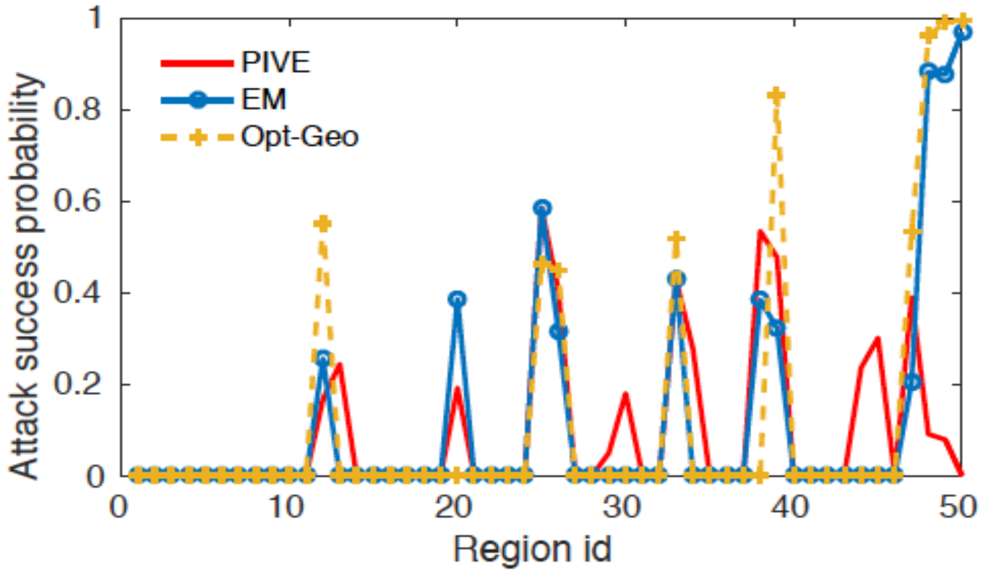
- Comparison with existing mechanisms on location privacy

EM - Laplacian-like mechanism uniform noise level



(a) Optimal inference attack

$$\hat{x} = \arg\min_{\hat{x} \in \mathcal{X}} \sum_{x \in \mathcal{X}} \Pr(x|x') d_p(\hat{x}, x)$$

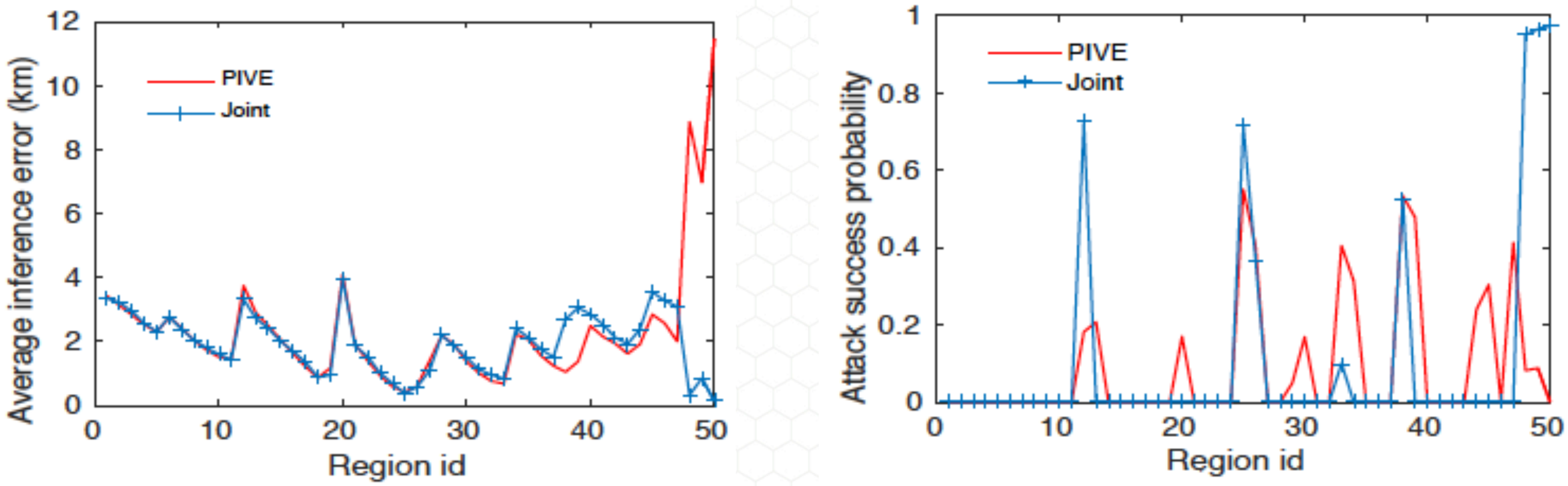


(b) Bayesian inference attack

$$\hat{x} = \arg\max_{x \in \mathcal{X}} \Pr(x|x')$$

Evaluation

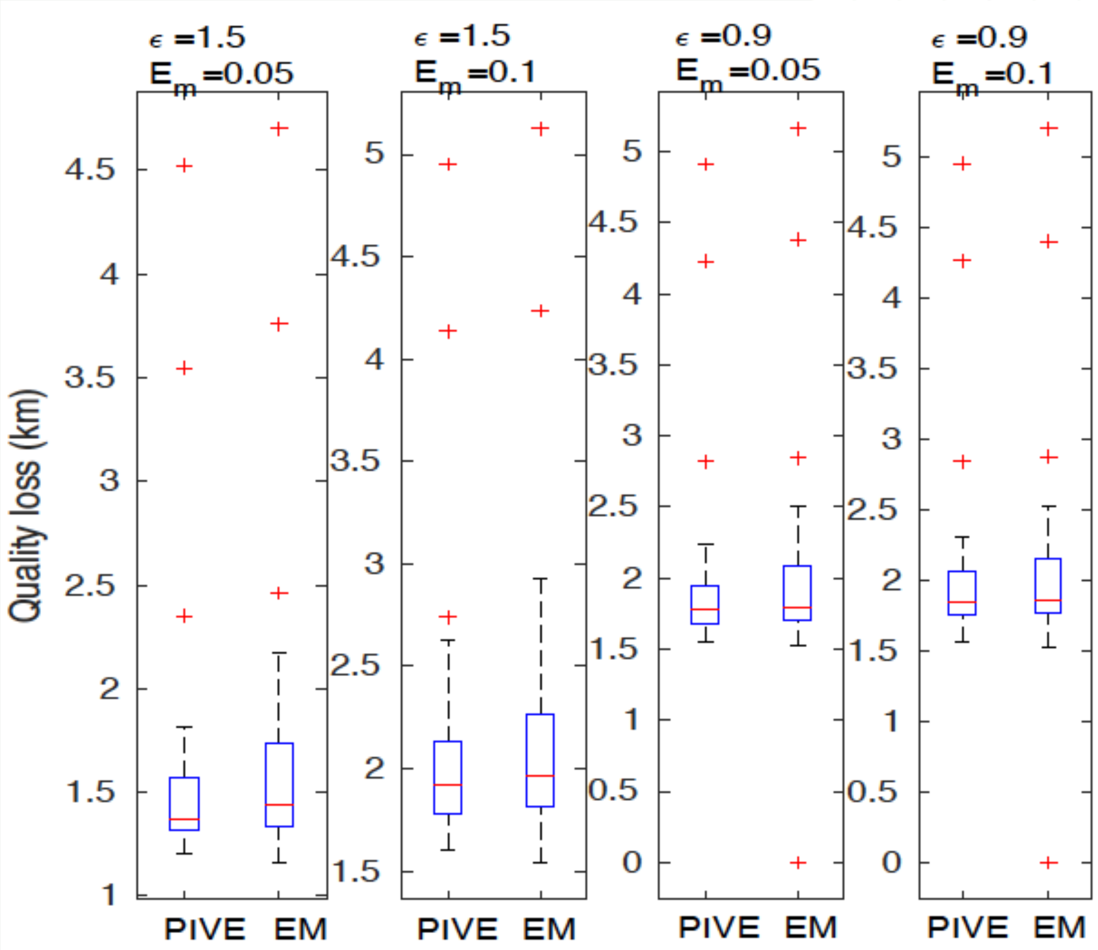
- Comparison with joint mechanism on location privacy



(a) Optimal inference attack

(b) Bayesian inference attack

Utility



Quality loss: the average distance between the fake location and the real location.

PIVE

- Geo-indistinguishability + prior information
- Adaptively adjust noise level of different privacy according to prior distribution
- Customizability

Thank you!

Q&A

Expected inference error

Conditional expected inference error

$$\sum_{\hat{x}, x \in \chi} \Pr(x|x') \, h(\hat{x}|x') \, d(\hat{x}, x)$$

the distance between the estimation and the actual location

Unconditional expected inference error

$$\sum_{x, x', \hat{x} \in \chi} \pi(x) f(x'|x) \, h(\hat{x}|x') \, d(\hat{x}, x)$$

Quality loss

$$\sum_{x, x' \in \chi} \pi(x) f(x'|x) \, d(x, x')$$

$h(\hat{x}|x')$ - Probability of guessing $\hat{x}$ as the user's actual location, given that x' is observed

Optimal inference attack: $\hat{x} = \operatorname{argmin}_{\hat{x} \in \chi} \sum_{x \in \chi} \Pr(x|x') \, d(\hat{x}, x)$

Bayesian inference attack: $\hat{x} = \operatorname{argmax}_{x \in \chi} \Pr(x|x')$